



Ogden, J., Summers, E., & Walker, S. (2021). *Patterns of Use: Conceptualising the role of web archives in online discourse*. Abstract from 4th RESAW (Research Infrastructure for the Study of Archived Web Materials) conference, Luxembourg.

Peer reviewed version

[Link to publication record in Explore Bristol Research](#)
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

PATTERNS OF USE

Conceptualising the role of web archives in
online discourse

Jessica Ogden

University of Bristol

jessica.ogden@bristol.ac.uk

@jessogden



Ed Summers

University of Maryland

edsu@umd.edu

@edsu

Shawn Walker

Arizona State University

shawn.w@asu.edu

@walkeroh

4th RESAW Conference - *Mainstream vs marginal content in Web history and Web archives* |
Luxembourg (Virtual) | 18 June 2021



INTRODUCTION

What's the use? Some background + 'use imaginaries'

01

THEORISING USE

An argument for theorising how web archives are used

02

OUTLINE

03

PATTERNS OF USE

Some categories of use, examples and discussion

04

CONCLUSIONS

Discussion and proposal for a research agenda on the 'social life' of web archives

01

INTRODUCTION

Some background on *The Social Life of Web Archives* project and a look at 'use imaginaries'



'USE IMAGINARIES' FOR WEB ARCHIVES

'Use Cases for Access to Internet Archive' (IIPC, 2006)

Types of Users

Professionals
Common citizens
Lawyers
Researchers

Types of Use

Journalistic documentation
'Digital management'
Employment documentation
Civil/Criminal Case Evidence
Patent Litigation
Genealogy

User Journeys

Searching for URL
Searching on Subject
Version/links comparison
Browsing Thematic Collections
'Personified Features'
Federated Search
Searching for Copies
Updates on Changes to Sites
Persistent References
On-demand harvesting
Data Mining...

'High Profile' Examples of Use

The Joy Reid controversy, from homophobic blog posts to a hacking claim, explained

MSNBC host Joy Reid was accused of writing homophobic blog posts. Reid at first said she was hacked — then apologized.

By German Lopez | @germanlopez | german.lopez@vox.com | Updated Apr 28, 2018, 8:09pm EDT



Bennett Rogin/WireImage via Getty Images

<https://www.vox.com/policy-and-politics/2018/4/27/17286392/joy-reid-msnbc-lgbtq-gay-hack>

Conservative party deletes archive of speeches from internet

Decade's worth of records is erased, including PM's speech praising internet for making more information available



▲ A speech in which David Cameron said the internet would help people hold politicians to account was among those deleted. Photograph: Barcroft Media

<https://www.theguardian.com/politics/2013/nov/13/conservative-party-archive-speeches-internet>

- Bush 'Mission Accomplished' (2003)
- UK Conservative Party speech deletions (Ballard, 2013; Ramesh and Hern, 2013)
- MH17 (Dewey, 2014; Taylor, 2014)
- Joy Ann Reid Blog (Nelson, 2018)

What's the Use?

'SCHOLARLY USE'

Overviews: (e.g.) Thomas et al., 2010; Stirling, Chevallier and Illien, 2012; Dougherty and Meyer, 2014; Hockx-Yu, 2014; Maemura, 2018

Community/Tool/Collection Development:
(e.g.) Lin et al., 2014; Jackson et al., 2016; Ruest et al. 2020

Surveys & Qualitative Research: (e.g.) Riley and Crookston, 2015; Costea, 2018; Vlassenroot et al., 2019;

BEYOND THE ACADEMY

User Needs & Behaviours: Costa and Silva, 2012; AlNoamany, Weigle and Nelson, 2013; AlNoamany et al., 2014

Legal/Copyright & IP:
Fagan, 2007; Eltgroth, 2009; Knutson, 2009; Quarles III and Crudo, 2014; McCarthy, 2018

Activist, Journalistic Fact-checking:
(Examples of use plenty, lacking analysis)

Social Media 'Misuse': (Acker & Chalet, 2020; Donovan, 2020; Littman, 2018; Zannettou et al., 2018)





02

THEORISING USE

Motivations for theorising how web archives are used

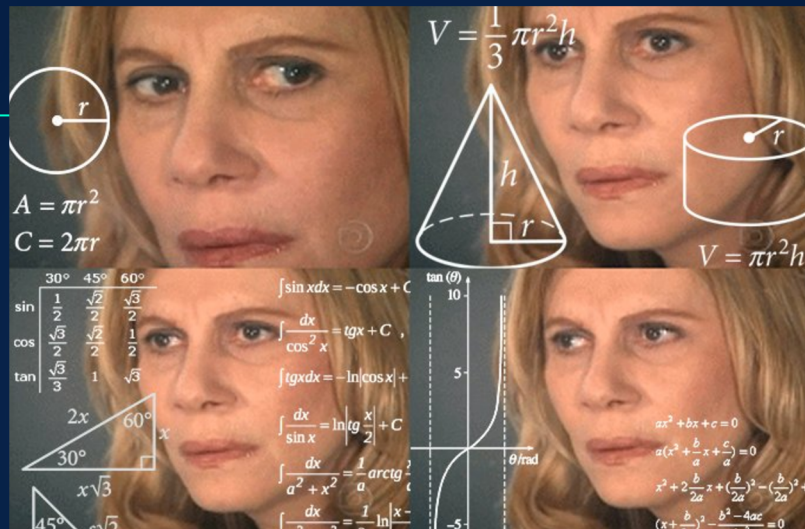
MOTIVATIONS

Some Questions

What are web archives **used for**?

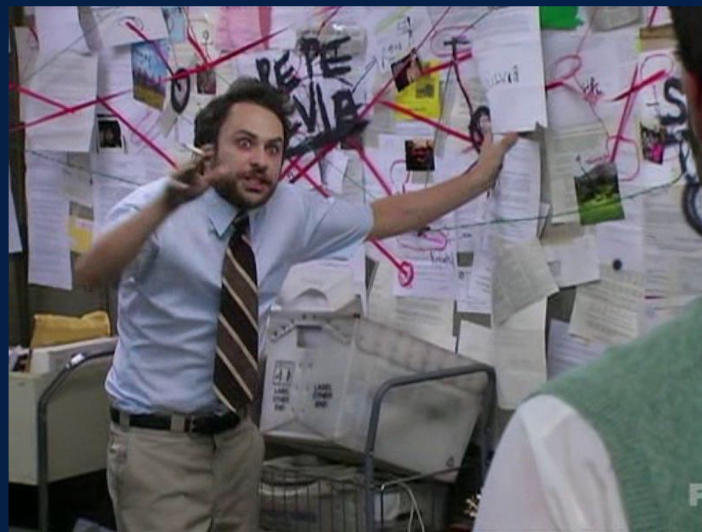
How might we **problematise** current boundaries and framings of web archival use?

What's the **significance** and **implications** of use for **a)** web archives themselves and **b)** our understanding of their role on the Web?



CONTRIBUTION OF THEORISING USE

- 1) **Explore** what constitutes use in the context of the archived Web + examine *situated* notions of 'misuse', disuse, non-use
- 2) Organise, consolidate and **mobilise** existing research on how web archives are used ("in the wild")
- 3) **Expand** and broaden notions of comes to count as a 'web archive'
- 4) **Connect** this research and the field to other relevant interdisciplinary internet studies work



IMPLICATIONS

For how archiving is done, terms and conditions of use

03

PATTERNS OF USE

Some categories of use, examples and discussion

PATTERNS OF USE (1)

Attribution

Ways that web archives provide mechanisms for users to **evidence, cite and attribute** online resources to particular actors/points in time


- **Linking** to and **embedding** content from a web archive snapshot
- Claims of **authenticity** and positioning of web archives as authoritative sources

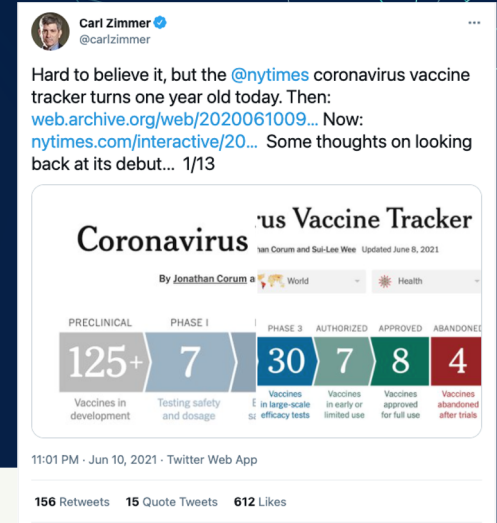
Examples

Perma.cc 

**Websites change.
Perma Links don't.**

Perma.cc helps scholars, journals, courts, and others create permanent records of the web sources they cite.

Perma.cc is supported by  **Cite safely.**



<https://twitter.com/carlzimmer/status/1403110178480869379>

PATTERNS OF USE (2)

Circumvention

Where web archives act as **alternative platforms** that enable users to (e.g.) avoid paywalls and/or social media content moderation by providing access to deleted or blocked content

- **Accessing, linking to and embedding** content and/or screenshots of archived content that is otherwise inaccessible
- Both **evasion** and preemptive tactics to enable accessibility of ephemeral content
- *Aside: Is all web archival use some kind of circumvention? (discuss!)*

Examples

archive.today
webpage capture




washingtonpost.com

search

search examples:

- washingtonpost.com for all snapshots from the host
- *.washingtonpost.com for list of subdomains
- <http://washingtonpost.com/> for exact url
- http://washingtonpost.com/* for url prefix

1..50 of 2627 urls →

Oldest	Newest	List of URLs, ordered from newer to older
	 18 Jun 2021 05:35	Record-setting heat blasts the West: 'Your skin is almost sizzling' - The Washington Post http://washingtonpost.com/climate-environment/2021/06/17/record-heat-wave-west/ redirected to • https://www.washingtonpost.com/climate-environment/2021/06/17/record-heat-wave-west/
	 18 Jun 2021 03:03	Jack Whittaker won Powerball's \$314 million jackpot. It ruined his life. - The Washington Post http://washingtonpost.com/history/2018/10/24/jack-whittaker-powerball-lottery-winners-life-was-ruined-after-m-jackpot/ redirected to • https://www.washingtonpost.com/history/2018/10/24/jack-whittaker-powerball-lottery-winners-life-was-ruined-after-m-jackpot/
	 18 Jun 2021 02:47	Powerball jackpot mystery: Who won \$731 million in Lonaconing, Md.? - The Washington Post http://washingtonpost.com/local/powerball-jackpot-maryland-lonaconing-winner/2021/06/16/501a958c-caed-11eb-81b1-34796c73933f_story.html redirected to • https://www.washingtonpost.com/local/powerball-jackpot-maryland-lonaconing-winner/2021/06/16/501a958c-caed-11eb-81b1-34796c73933f_story.html

PATTERNS OF USE (2)

Circumvention

Where web archives act as **alternative platforms** that enable users to (e.g.) avoid paywalls and/or social media content moderation by providing access to deleted or blocked content

- **Accessing, linking to and embedding** content and/or screenshots of archived content that is otherwise inaccessible
- Both **evasion** and preemptive tactics to enable accessibility of ephemeral content
- *Aside: Is all web archival use some kind of circumvention? (discuss!)*

Examples

Humans and technology

Covid hoaxes are using a loophole to stay alive— even after content is deleted

Pandemic conspiracy theorists are using the Wayback Machine to promote “zombie content” that evades moderators and fact-checkers.

by **Joan Donovan**

April 30, 2020

<https://www.technologyreview.com/2020/04/30/1000881/covid-hoaxes-zombie-content-wayback-machine-disinformation/>

PATTERNS OF USE (2)

Circumvention

Where web archives act as **alternative platforms** that enable users to (e.g.) avoid paywalls and/or social media content moderation by providing access to deleted or blocked content

- **Accessing, linking to and embedding** content and/or screenshots of archived content that is otherwise inaccessible
- Both **evasion** and preemptive tactics to enable accessibility of ephemeral content
- *Aside: Is all web archival use some kind of circumvention? (discuss!)*

Examples

Humans and technology

Covid hoaxes are using a ve— is

SEPTEMBER 28, 2020

BY AMELIA ACKER

School of Information, The University of
Texas at Austin, USA

MITCH CHAIET

Moody College of Communication, The
University of Texas at Austin, USA

PEER REVIEWED

The weaponization of web archives: Data craft and COVID-19 publics

An unprecedented volume of harmful health misinformation linked to the coronavirus pandemic has led to the appearance of misinformation tactics that leverage web archives in order to evade content moderation on social media platforms. Here we present newly identified manipulation techniques designed to maximize the value, longevity, and spread of harmful and non-factual content across social media using provenance information from web archives and social media analytics. After identifying conspiracy content that has been archived by human actors with the Wayback Machine, we report on user patterns of “screensampling,” where images of archived misinformation are spread via social platforms. We argue that archived web resources from the Internet Archive’s Wayback Machine and

ing the Wayback Machine to
s moderators and fact-

April 30, 2020

<https://misinforeview.hks.harvard.edu/article/the-weaponization-of-web-archives-data-craft-and-covid-19-publics/>

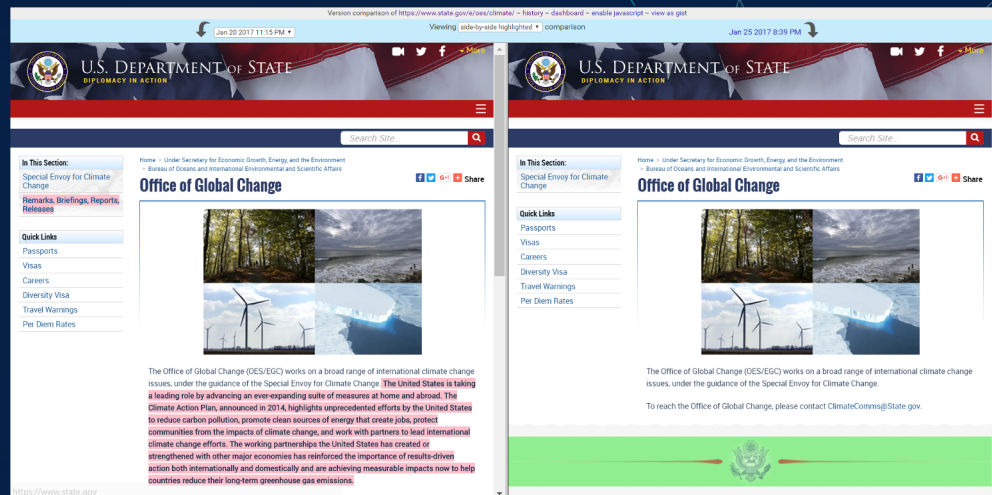
PATTERNS OF USE (3)

Monitoring 📡📡

How web archives provide the data and tools for **observing online actors and change** over time, and the potential role that preservation plays in mechanising different forms of 'watching'

- **Snapshot** collection approach provides data over time
- Standard 'diffing' tools enable semi-automated **change detection** in source code
- **Change, addition or removal** of links, navigation, content, language

Examples



Environmental Data & Governance Initiative Web Monitoring

Further info: *Access Denied: Federal Web Governance Under the Trump Administration* (Gehrke et al., 2021)

PATTERNS OF USE (3)

Monitoring 📡📡

How web archives provide the data and tools for **observing online actors and change** over time, and the potential role that preservation plays in mechanising different forms of 'watching'

- **Snapshot** collection approach provides data over time
- Standard 'diffing' tools enable semi-automated **change detection** in source code
- **Change, addition or removal** of links, navigation, content, language

Examples



Platform Governance Archive 1.0 : <https://pga.hiig.de/>

PATTERNS OF USE (4)

“Computation”/Data Mining

How web archives are **consumed by one-way computational processes** that rely on the extraction and transformation of archived content

- Training **language models** for text generation, summarization, translation.
- Training **neural network models** for image/video recognition, recommendation, classification and segmentation.
- **Automation, simulation, surveillance, prediction.**

Examples

OpenAI
spaCy



EleutherAI

IMAGENET

amazon Rekognition



Common Crawl



04

CONCLUSIONS

Discussion, conclusions and an agenda for future work

DISCUSSION, CONCLUSIONS & FUTURE WORK

- Web archival use remains **under-examined** and **under-theorised**
- We began collating, **mapping patterns** and categories of web archival use through a conceptual framework
- Brief examples highlight how web archives become **embedded in everyday practice** online, but there are more
- **Complicates** the role(s) and broader significance of web archives (beyond *just* tools for research into the Web's past) - implications, risks
- **Much more to be done!** Creative methods and interdisciplinary perspectives required



Work in progress: Using web archives to study web archives!

THANKS. Questions?

Jessica Ogden

University of Bristol



jessica.ogden@bristol.ac.uk



@jessogden

This work was supported by the UK Economic and Social Research Council (Grant No. ES/V012177/1).



Economic
and Social
Research Council

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.

REFERENCES (1)

Acker, A. and Chalet, M. (2020) 'The weaponization of web archives: Data craft and COVID-19 publics', Harvard Kennedy School Misinformation Review, 1(3). doi: 10.37016/mr-2020-41.

AlNoamany, Y. et al. (2014) 'Who and what links to the Internet Archive', International Journal on Digital Libraries, 14(3), pp. 101–115. doi: 10.1007/s00799-014-0111-5.

AlNoamany, Y., Weigle, M. C. and Nelson, M. L. (2013) 'Access Patterns for Robots and Humans in Web Archives', arXiv:1309.4009 [cs]. Available at: <http://arxiv.org/abs/1309.4009> (Accessed: 14 June 2019).

Ballard, M. (2013) Conservatives erase Internet history, Computer Weekly. Available at: <https://www.computerweekly.com/blog/Public-Sector-IT/Conservatives-erase-Internet-history> (Accessed: 13 March 2019).

Costa, M. and Silva, M. J. (2012) 'Evaluating Web Archive Search Systems', in Wang, X. S. et al. (eds) Web Information Systems Engineering - WISE 2012: 13th International Conference, Paphos, Cyprus, November 28-30, 2012. Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 440–454. Available at: http://dx.doi.org/10.1007/978-3-642-35063-4_32.

Costea, M.-D. (2018) Report on Scholarly Use of Web Archives. Aarhus, Denmark: NetLab, pp. 1–27. Available at: http://netlab.dk/wp-content/uploads/2018/02/Costea_Report_on_the_Scholarly_Use_of_Web_Archives.pdf (Accessed: 27 February 2018).

Dewey, C. (2014) 'How Web archivists and other digital sleuths are unraveling the mystery of MH17', Washington Post, 21 July. Available at: <https://www.washingtonpost.com/news/the-intersect/wp/2014/07/21/how-web-archivists-and-other-digital-sleuths-are-unraveling-the-mystery-of-mh17/> (Accessed: 10 July 2019).

Donovan, J. (2020) Covid hoaxes are using a loophole to stay alive—even after content is deleted, MIT Technology Review. Available at: <https://www.technologyreview.com/2020/04/30/1000881/covid-hoaxes-zombie-content-wayback-machine-disinformation/> (Accessed: 28 June 2021).

REFERENCES (2)

- Dougherty, M. and Meyer, E. T. (2014) 'Community, Tools, and Practices in Web Archiving: The State-of-the-Art in Relation to Social Science and Humanities Research Needs', *Journal of the Association for Information Science and Technology*, 65(11), pp. 2195–2209.
- Eltgroth, D. R. (2009) 'Best Evidence and the Wayback Machine: Toward a Workable Authentication Standard for Archived Internet Evidence', *Fordham Law Review*, 78(1), pp. 181–215. Available at: <https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=4467&context=flr> (Accessed: 7 February 2019).
- Fagan, M. (2007) "Can you do a Wayback on that?" The Legal Community's Use of Cached Web Pages In and Out of Trial', *Boston University Journal of Science & Technology Law*, 13(1). Available at: https://www.bu.edu/law/journals-archive/scitech/volume131/documents/fagan_web.pdf (Accessed: 27 May 2021).
- Gehrke, G. et al. (2021) Access Denied: Federal Web Governance Under the Trump Administration. Environmental Data & Governance Initiative, p. 38.
- Hockx-Yu, H. (2014) 'Access and Scholarly Use of Web Archives', *Alexandria: The Journal of National and International Library and Information Issues*, 25(1–2), pp. 113–127. doi: 10.7227/ALX.0023.
- IIPC Access Working Group (2006) Use Cases for Access to Internet Archive. International Internet Preservation Consortium, p. 28. Available at: http://web.archive.org/web/20060613164905id_/https://netpreserve.org/publications/iipc-r-003.pdf (Accessed: 16 June 2021).
- Jackson, A. et al. (2016) 'Desiderata for Exploratory Search Interfaces to Web Archives in Support of Scholarly Activities', in *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries*. New York, NY, USA: ACM (JCDL '16), pp. 103–106. doi: 10.1145/2910896.2910912.
- Knutson, A. N. (2009) 'Proceed With Caution: How Digital Archives Have Been Left in the Dark', *Berkeley Technology Law Journal*, 24(1), pp. 437–473. Available at: <https://www.jstor.org/stable/24121350> (Accessed: 27 May 2021).

REFERENCES (3)

Lin, J., Kraus, K. and Punzalan, R. L. (2014) 'Supporting "Distant Reading" for Web Archives', in. Digital Humanities, Lausanne, Switzerland, pp. 239–241. Available at: http://users.umi.acs.umd.edu/~jimmylin/publications/Lin_etal_DH2014.pdf (Accessed: 26 September 2018).

Littman, J. (2018) Islamic State Extremists Are Using the Internet Archive to Deliver Propaganda, Medium. Available at: <https://medium.com/on-archivy/islamic-state-extremists-are-using-the-internet-archive-to-deliver-propaganda-a132597dd16> (Accessed: 12 November 2019).

Maemura, E. (2018) 'What's Cached is Prologue: Reviewing Recent Web Archives Research Towards Supporting Scholarly Use', in Proceedings of the ASIS&T Annual Meeting 2018. Annual Meeting of the Association for Science & Technology, Vancouver, BC. Available at: <http://hdl.handle.net/1807/89426> (Accessed: 21 July 2018).

McCarthy, K. (2018) 'Archive.org's Wayback Machine is legit legal evidence, US appeals court judges rule', The Register, 4 September. Available at: https://www.theregister.co.uk/2018/09/04/wayback_machine_legit/ (Accessed: 30 January 2019).

Nelson, M. L. (2018) '2018-04-24: Why we need multiple web archives: the case of blog.reidreport.com', Web Science and Digital Libraries Research Group, 24 April. Available at: <https://ws-dl.blogspot.com/2018/04/2018-04-24-why-we-need-multiple-web.html> (Accessed: 2 June 2021).

Quarles III, J. L. and Crudo, R. A. (2014) '[Way]Back to the Future: Using the Wayback Machine in Patent Litigation', Landslide, 6(3). Available at: https://www.americanbar.org/groups/intellectual_property_law/publications/landslide/2013-14/january-february/wayback-future/.

Ramesh, R. and Hern, A. (2013) 'Conservative party deletes archive of speeches from internet', The Guardian, 13 November. Available at: <https://www.theguardian.com/politics/2013/nov/13/conservative-party-archive-speeches-internet> (Accessed: 10 July 2019).

Riley, H. and Crookston, M. (2015) Awareness and use of the New Zealand Web Archive: A survey of New Zealand academics. New Zealand: National Library of New Zealand and the University of Wellington Victoria, p. 44. Available at: <https://natlib.govt.nz/files/webarchive/nzwebarchive-awarenessanduse.pdf> (Accessed: 25 May 2021).

REFERENCES (4)

Ruest, N. et al. (2020) 'The Archives Unleashed Project: Technology, Process, and Community to Improve Scholarly Access to Web Archives', in JCDL'20: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020. ACM/IEEE Joint Conference on Digital Libraries 2020, Wuhan, Hubei, China: ACM, pp. 157–166. Available at: <https://doi.org/10.1145/3383583.3398513>.

Stirling, P., Chevallier, P. and Illien, G. (2012) 'Web Archives for Researchers: Representations, Expectations and Potential Uses', D-Lib Magazine, 18(3/4). Available at: <http://www.dlib.org/dlib/march12/stirling/03stirling.html> (Accessed: 1 February 2016).

Taylor, N. (2014) The MH17 Crash and Selective Web Archiving, The Signal. Available at: <https://blogs.loc.gov/thesignal/2014/07/21503/> (Accessed: 8 March 2019).

Thomas, A. et al. (2010) Researcher Engagement with Web Archives: Challenges and Opportunities for Investment. Joint Information Systems Committee Report. Available at: <https://ssrn.com/abstract=1715000> (Accessed: 8 July 2018).

Vlassenroot, E. et al. (2019) 'Web archives as a data resource for digital scholars', International Journal of Digital Humanities, 1, pp. 85–111 doi: 10.1007/s42803-019-00007-7. Available at: <https://doi.org/10.1007/s42803-019-00007-7>

Zannettou, S. et al. (2018) 'Understanding Web Archiving Services and Their (Mis)Use on Social Media', in Proceedings of the 12th International AAAI Conference on Web and Social Media. AAAI Conference on Web and Social Media (ICWSM), Palo Alto, CA: arxiv. Available at: <https://arxiv.org/abs/1801.10396> (Accessed: 9 July 2019).

